1M4-2

信頼度つきギャップ分析による社会ネットワークからの高中心性ノード群同定

Identifying High Centrality Nodes in Social Network based on Gap Analysis with Confidence Level

大原剛三 *1 斉藤和巳 *2 木村昌弘 *3 元田浩 *4
Kouzou Ohara Kazumi Saito Masahiro Kimura Hiroshi Motoda

*1青山学院大学 Aoyama Gakuin University
*2静岡県立大学 University of Shizuoka
*3龍谷大学 Ryukoku University
*4大阪大学 Osaka University

This paper addresses a problem of identifying nodes having a high centrality value in a large social network based on its approximation derived only from nodes sampled from the network. We assume that a gap exists between two adjacent nodes ordered in descending order of approximations of true centrality values if it can divide the ordered list of nodes into two groups so that any node in one group has a higher centrality value than any one in another group with a given confidence level. Then, we incorporate confidence intervals of true centrality values and devise an efficient algorithm that applies a resampling-based framework to estimate the intervals as accurately as possible. Using a real world large social network, we empirically show that the gaps detected by the proposed method enable us to correctly identify a set of nodes having a high centrality value.

1. はじめに

今日，Facebook や Twitter などのソーシャルメディアの普及により，インターネット上には巨大な社会ネットワークが構築されている．ソーシャルメディアに一旦投稿された情報は，そのような社会ネットワークを通して急速，かつ広範囲に拡散され，我々の日常における意思決定にも多大な影響を与えるため，近年，社会学のみならず計算機科学も含めた多様な分野において社会ネットワークの分析が進められている [Kleinberg 08, Chen 13].

そのような社会ネットワーク分析においては，幾つかの中心性と呼ばれる指標が利用されている [Katz 53, Freeman 79, Bonacichi 87, Brin 98, Zhuge 10]．中心性はネットワーク構造に基づきノードを特徴づけるものであり，その値から各ノードがどのような意味で，どの程度重要かについての情報を我々にもたらしてくれる．また，ネットワークのスケールフリー性が次数分布から導かれるように，ネットワーク全体の構造的特徴を知る手がかりともなる．一方，近接中心性や媒介中心性などのように，その値を求めるために対象ノードの隣接ノードの情報のみならず，任意のノード間の最短経路などのようなネットワーク全体にわたる情報を必要とするものがあり，それらに関しては，ネットワークが大きくなるとその計算が困難になる．

実際には，そのような計算コストの高い中心性は，ノードペアなどから導かれる値を基礎に，その平均値として定義されることが多い．このことから，その計算コスト軽減に対する 1 つのアプローチとしては，サンプリングによるノード数の削減が考えられる．ノード数を制限することにより中心性の計算は容易になるが，得られるのは近似値となるため，真の値との近似誤差を精度良く推定することが重要となる．この問題に対して，我々は近似誤差を精度よく推定するリサンプリング法に基づいた枠組みを提案し，それにより得られる近似誤差（以下，リサンプリング誤差）が独立同分布の下でのサンプリングを仮定した標準的な近似誤差（以下，標準誤差）よりも正確な誤差範囲を与えることを実験的に示している [Ohara 14].

一方，社会ネットワーク分析では，全ノードの中心性の値を知ることよりも，高い中心性をもつノードが興味の対象となることが多い．そのため本稿では，高い中心性をもつノード集合をサンプリングにより得られた中心性の近似値から精度よく同定することを考える．具体的には，ノードを中心性の近似値の降順に並べ，あるノード間で 2 つに分割したとき，上位集合中の任意のノードが下位集合中のどのノードよりも大きい真の中心性の値をもつ場合，それらのノード間にはギャップがあるとし，そのようなギャップを中心性の近似値のみから一定の精度で検出することを試みる．統計的な観点からは，これは，与えられた信頼度の下で各ノードの中心性指標値の信頼区間を求め，分割後の上位集合，下位集合間のその重複関係を調べることに相当する．そこで本研究では，その信頼区間の導出に前述のリサンプリング誤差を導入し，実際の大規模社会ネットワークを用いた評価実験を通して，標準誤差を利用するよりも多くのギャップを検出し，かつ検出したギャップが高い中心性をもつノード集合の同定に有用であることを示す．

連絡先: 大原剛三，青山学院大学理工学部情報テクノロジー学科，〒 252-5258 相模原市中央区淵野辺 5-10-1，ohara@it.aoyama.ac.jp

2. リサンプリング法に基づいた近似誤差推定

本節では，文献 [Ohara 14] に従い，リサンプリング法に基づいた近似誤差推定の一般的な枠組み，およびその近接中心性と媒介中心性への適用について述べる．

2.1 一般的枠組み

いま，ある集合 S（$|S| = L$）に対して，f を S 中の各要素に何らかの値を対応付ける関数とする．このとき，S に対する f の平均値 $\mu = (1/L)\sum_{s\in S} f(s)$ を，S の任意の部分集合 T（$|T| = N$）に対する f の値 $\{f(t)|t \in T, T \subset S\}$ のみから推定することを考える．実際には，T に対する f の値から μ を直接推定することはできないため，μ と $\mu(T) = (1/N)\sum_{t\in T} f(t)$ 間の近似誤差を，μ を仮定せずに推定する．そのために，任意の $T \in \mathcal{T}$ に対して，$T \subset S$，かつ $|T| = N$ であるような S の部分集合族 $\mathcal{T} \subset 2^S$ を考える．このとき，μ と $\mu(T)$ の近似誤差 $RE(N)$ を以下のように定義する．

$$RE(N) = \sqrt{\langle(\mu - \mu(T))^2\rangle_{T\in\mathcal{T}}} = \sqrt{\frac{L-N}{(L-1)N}} \times \sqrt{\frac{1}{L}\sum_{s\in S}(f(s)-\mu)^2} \quad (1)$$

この式は，S から N 個の要素をリサンプリングすることで得られる $\mathcal{T}$ に対して，$T \in \mathcal{T}$ に対する部分平均 $\mu(T)$ と真の平均 μ との二乗平均平方根誤差（RMSE）を計算していると解釈できる．ここで，右辺のうち N に依存するのは第 1 項のみであり，第 2 項は N に依存しないこと，および，この第 2 項が全体集合 S に対する関数 f の値の標準偏差となっていることから，第 2 項を定数項 σ，第 1 項をその係数項 $C(N)$ とし，$RE(N) = C(N)\sigma$ とする．このことから，実際には部分集合 T をリサンプリングせず，定数 L，σ，およびサンプリング数 N が与えられた時点で $RE(N)$ の値を計算可能なことがわかる．以下では，この $RE(N)$ をリサンプリング誤差と呼ぶ．

一方，より一般には，独立同分布の下でのサンプリングを前提に，μ と $\mu(T)$ の近似誤差の期待値を計算する．具体的には，$t \in T$ がある確率分布 $p(t)$ に従って独立に S から選択されたと仮定する．$p(t)$ としては，$p(t) = 1/L$ のような経験的な一様分布などが考えられる．このとき，μ と $\mu(T)$ の近似誤差の期待値は次式のように定義できる．

$$SE(N) = \sqrt{\langle(\mu - \mu(T))^2\rangle} = \sqrt{\frac{1}{N}} \times \sqrt{\frac{1}{L}\sum_{s\in S}(f(s) - \mu)^2} \qquad (2)$$

この式も式 (1) 同様，右辺の第 1 項のみが N に依存し，第 2 項は関数 f の値の標準偏差となっていることから，実際には T をサンプリングすることなく，その値を求めることができる [*1]．以下，$SE(N)$ を標準誤差と呼び，式 (1) 同様，右辺の第 2 項を定数項 σ，第 1 項をその係数項 $D(N)$ とし，$SE(N) = D(N)\sigma$ とする．ここで，$C(N) \leq D(N)$ であり，$C(L) = 0$ であるのに対し $D(L) \neq 0$ であることに注意されたい．すなわち，ある N に対して $RE(N) \leq SE(N)$ であり，$N = L$ のとき $RE(N)$ は 0 となるが，$SE(N)$ は 0 とはならない．

2.2 中心性指標への適用

次に，上記の近似誤差推定の枠組みを社会ネットワークにおけるノード中心性の推定問題に適用する．以下では，社会ネットワークを有向グラフ $G = (V, E)$ により表現する．ここで，V，および $E \subseteq V \times V$ はそれぞれネットワーク中のノード集合と有向リンク集合である．

2.2.1 近接中心性

まず，G 中のノード $u \in V$ に対して次式で定義される近接中心性を考える．

$$cls_G(u) = \frac{1}{(|V| - 1)} \sum_{v\in V, v\neq u} \frac{1}{spl_G(u, v)} \qquad (3)$$

ここで，$spl_G(u, v)$ はグラフ G におけるノード u からノード v までの最短経路長を表し，v が u から到達可能でなければ $spl_G(u, v) = \infty$ とする．直観的には，ネットワーク中の他のどのノードにも比較的短い経路長で到達可能なノードほど近接中心性は大きな値となる．この近接中心性を計算する一般的な方法としては，基点ノードから 1 つのリンクを辿ることで新たに到達可能となるノード集合を漸進的に求める burning アルゴリズム [Newman 01] が知られているが，各ノード u に対する近接中心性 $cls_G(u)$ を求める計算量は $O(|E|)$ であり，巨大な社会ネットワークに対しては膨大な計算時間を要する．

この近接中心性に対して，前述のリサンプリングに基づいた近似誤差推定の枠組みを適用することを考える．ここでは，対象ノード u を除く V からサンプリングしたノード集合 T（$|T| = N$）のみから求められる u の近接中心性の近似値 $cls_G(u; T)$ と真の値 $cls_G(u)$ の近似誤差を考えることになる．そのために，前節における全サンプル集合 S，評価関数 f を近接中心性の計算に合わせて具体化する．まず，近接中心性はノード集合全体に対する値ではなく，各ノードに対する値であるため，S に関しては，対象ノードを u としたとき，$S_u = V \setminus \{u\}$ とする．ここで，$\setminus$ は集合差を意味する．一方，$cls_G(u)$ はその定義より，ノード u 以外のノード v に対して求められる $1/spl_G(u, v)$ の平均値であるため，評価関数 f に関しては，$f_u(v) = 1/spl_G(u, v)$ とする．これにより，$cls_G(u; T)$ を $(1/N)\sum_{v\in T} f_u(v)$ として求めることができ，式 (1)，および (2) に従い，$RE(N)$，$SE(N)$ をそれぞれ計算することが可能となる．

2.2.2 媒介中心性

次に，次式で定義されるノード u の媒介中心性について考える．

$$btw_G(u) = \frac{1}{(|V| - 1)(|V| - 2)} \sum_{v\in V, v\neq u} \left(\sum_{\substack{w\in V, w\neq u \\ w\neq v}} \frac{nsp_G(v, w; u)}{nsp_G(v, w)} \right) \qquad (4)$$

ここで，$nsp_G(v, w)$ はグラフ G におけるノード v から w までの最短経路数，$nsp_G(v, w; u)$ はそのうちノード u を経由する最短経路数を表す．直観的には，ノード u を経由する 2 ノード間の最短経路数が多いほど，u の媒介中心性 $btw_G(u)$ の値は大きくなる．この媒介中心性を求める標準的な方法としては，Brandes のアルゴリズム [Brandes 01] が知られており，各ノード u に対して $btw_G(u)$ を求める計算量は近接中心性同様 $O(|E|)$ である．

いま，ノード u の真の媒介中心性の値 $btw_G(u)$ と，u を除く V の部分集合 T（$|T| = N$）から求められるその近似値 $btw_G(u; T)$ の近似誤差を 2.1 節の枠組みに基づき推定することを考える．全サンプル集合 S に関しては，媒介中心性も全ノード集合ではなく個々のノードに対して定まる値であるため，近接中心性と同様に対象ノード u に対して $S_u = V \setminus \{u\}$ とする．一方，式 (4) 中のカッコ内の項を関数 $btw_G(u; v)$ とすると，$btw_G(u)$ はノード u 以外のノード v に対して求められる $btw_G(u; v)/(|V| - 2)$ の平均と考えられる．したがって，$f_u(v) = btw_G(u; v)/(|V| - 2)$ とすることで，任意の部分集合 T に対するノード u の媒介中心性 $btw_G(u; T)$ を $(1/N)\sum_{v\in T} f_u(v)$ として求めることができ，式 (1)，(2) に従い $RE(N)$，$SE(N)$ をそれぞれ計算することができる．

3. 信頼度つきギャップ検出法

本節では，ネットワーク中のノードの部分集合から推定される中心性の近似値のみを用いて，与えられた信頼度の下で実際に高い中心性をもつノード集合を同定する手法を考える．まず，ここでの問題を形式的に定義する．ネットワーク $G(V, E)$ に対して，$\mu_G(v)$ をノード $v \in V$ の真の中心性指標値とし，$\mu_G(v; T)$ をノードの部分集合 $T \subseteq V$ から得られるその近似値，$\sigma(v; |T|)$ を $RE(v; |T|)$ や $SE(v; |T|)$ のような近似誤差とする．また，ノード v が与えられたとき，$\mu_G(v; T)$ に基づくノード集合 V の互いに疎な分割 $V_H(v; T) = \{u \in V; \mu_G(u; T) \geq \mu_G(v; T)\}$，および $V_L(v; T) = \{w \in V; \mu_G(w; T) < \mu_G(v; T)\}$ を考える．このとき，統計における信頼区間推定の考えの下，ここでの問題は，任意の $u \in V_H(v; T)$ と $w \in V_L(v; T)$ が以下の不等式を満たすようなノード $v \in V$ をすべて見つける問題と定義できる．

$$\mu_G(u; T) - z(\alpha) \cdot \sigma(u; |T|) > \mu_G(w; T) + z(\alpha) \cdot \sigma(w; |T|) \qquad (5)$$

*1 $RE(N)$，$SE(N)$ いずれの計算においても σ が必要となるが，$|S| = L$ が大きい場合はそもそも σ の計算が困難であるため，実際にはその近似値として，$|S'| = L'$ が十分小さい部分集合 $S' \subset S$ から現実的な計算時間で得られる標準偏差 σ' を近似値として用いる．

ここで，$0 < \alpha < 1$ であり，$z(\alpha)$ は標準正規分布における信頼度 $C = 100(1-\alpha)\%$ に対する上側信頼限界値である．言い換えるなら，この不等式を満たす場合，信頼度 C で任意の $u \in V_H(v;T)$ と $w \in V_L(v;T)$ に対して $\mu_G(u) > \mu_G(w)$ が成り立つ．ここで，上位集合 $V_H(v;T)$ が我々が同定したいノード集合であり，以下，ノード v と $v' \in \arg\max_{w \in V_L(v;T)} \mu_G(w;T)$ 間にはギャップが存在するという．この問題をナイーブに解く場合，各ノード v に対して $|V_H(v;T)||V_L(v;T)|$ 個のノードペアについて上記の不等式を満たすかどうかを調べる必要があるため，その計算量は $O(|V|^3)$ となり，ネットワークが大規模化した場合はその計算は困難となる．

これに対して，$V_H(v;T)$ の誤差下限 $\min_{u \in V_H(v)}(\mu_G(u;T) - z(\alpha)\sigma(u;|T|))$，および $V_L(v;T)$ の誤差上限 $\max_{w \in V_L(v)}(\mu_G(w;T) + z(\alpha)\sigma(w;|T|))$ をそれぞれ $LB(V_H(v);T,\alpha)$ と $UB(V_L(v);T,\alpha)$ したとき，ここでの問題は，与えられた α に対して $LB(V_H(v);T,\alpha) > UB(V_L(v);T,\alpha)$ を満たすすべての $v \in V$ を見つける問題と考えられる．$LB(V_H(v);T,\alpha)$ と $UB(V_L(v);T,\alpha)$ は，ノード集合 V を一度走査するだけで任意の $v \in V$ に対して同時に計算可能であるため，全体の計算量は $O(|V|^2)$ となる．しかし，ネットワークが大きくなった場合，そのようなノードをすべて見つけることはまだ難しい．

そこで，本研究では，中心性指標の近似値 $\mu_G(v;T)$ の降順に並べたノードリスト $(v_1, v_2, \cdots, v_{|V|})$ を考える．すなわち，任意の $i \in \{1, \cdots, |V|-1\}$ に対して，$\mu_G(v_i;T) \geq \mu_G(v_{i+1};T)$ を仮定する．このとき，$LB(V_H(v_k);T,\alpha)$ は $LB(V_H(v_k);T,\alpha) = \min(LB(V_H(v_{k-1});T,\alpha),\ \mu_G(v_k;T) - z(\alpha)\sigma(v_k;|T|))$ というように再帰的に定義可能であり，同様に，$UB(V_L(v_k);T,\alpha)$ も $UB(V_L(v_k);T,\alpha) = \max(UB(V_L(v_{k+1});T,\alpha),\ \mu_G(v_{k+1};T) + z(\alpha)\sigma(v_{k+1};|T|))$ と定義できる．この定義に従えば，すべての $v \in V$ に対する $LB(V_H(v);T,\alpha)$ と $UB(V_L(v);T,\alpha)$ をノードリスト $(v_1, v_2, \cdots, v_{|V|})$ をそれぞれに対して一度走査するだけで求めることが可能となる．これは，ノードリストを二度走査するだけですべてのギャップを同定可能であることを意味する．具体的には，最初の走査（forward ステップ）において，k を 1 から $|V|-1$ まで変化させつつ $LB(V_H(v_k);T,\alpha)$ を求め，続く二度目の走査（backward ステップ）において，k を $|V|$ から 2 まで変化させつつ $UB(V_L(v_k);T,\alpha)$ を計算し，$LB(V_H(v_k);T,\alpha) > UB(V_L(v_k);T,\alpha)$ が成り立つ場合にギャップを同定する．この手法における計算量に関しては，ノード集合のソーティングにかかる計算量が支配的となるため，$O(|V|\log|V|)$ と考えることができ，大規模な社会ネットワークに対しても現実的な時間でのギャップ分析が可能といえる．以下に，同定したギャップの集合を A としたときの提案法の手続きをまとめる．

1. （初期化）
   $A \leftarrow \emptyset$，$LB(V_H(v_1);T,\alpha) = \mu_G(v_1;T) - z(\alpha)\sigma(v_1;|T|))$，$UB(V_L(v_{|V|});T,\alpha) = 0$ とする．
2. （Forward step）
   k を 2 から $|V|-1$ まで変化させ，$LB(V_H(v_k);T,\alpha)$ を再帰的に計算．
3. （Backward step）
   k を $|V|-1$ から 2 まで変化させ，以下を実行．
   ・$UB(V_L(v_k);T,\alpha)$ を再帰的に計算．
   ・$A \leftarrow A \cup \{v_k\}$ if $LB(V_H(v_k);T,\alpha) > UB(V_L(v_k);T,\alpha)$
4. 解集合 A を出力して終了．

以下では，式 (5) において，近似誤差 $\sigma(v;|T|)$ として $\sigma(v;|T|) = 0$，$\sigma(v;|T|) = SE(v;|T|)$，$\sigma(v;|T|) = RE(v;|T|)$ を用いた 3 つの手法を考え，それぞれナイーブ法，SE 法，RE 法と呼ぶ．ナイーブ法は常に $\mu_G(v;T) = \mu_G(v)$ を仮定するため，$\mu_G(v_k;T) \neq \mu_G(v_{k+1};T)$ であるようなすべての k に対して，ノード v_k と v_{k+1} の間にはギャップが存在すると同定する．一方，$RE(v;|T|)$ と比較して，$SE(v;|T|)$ は $\mu_G(v;T)$ の近似誤差を過大評価するため，SE 法により同定されるギャップ数は RE 法に比べて少なくなる．次節では，これらの手法を実世界の社会ネットワークを用いて実験的に評価する．

4. 評価実験

前節で提案したギャップ検出法を，実際の大規模ネットワークを用いて実験的に評価した．本実験で用いたネットワークは，Twitter[*2] から抽出したフォロワーネットワークである．具体的には，2011 年 3 月 5 日から 3 月 24 日までの約 3 週間にわたって収集した 201,297,161 ツイートの投稿者から，この期間中に 200 件以上のツイートをした 1,088,040 名の投稿者を抽出し，その投稿者間のフォロー関係をネットワーク化した有向グラフを作成した．そのノード数は 1,088,040，リンク数は 157,371,628 である．

本実験では，このネットワーク中のノードのうち，すべてのノードから求めた真の中心性の値において上位 100 ノードを対象にナイーブ法，SE 法，RE 法の各手法を評価した．実験手順としては，ノード集合 V からノードを 1 つずつランダム非復元抽出し，それを順次部分集合 T に加え，ノード被覆率 $|T|/|V|$ が 0.01 増加するごとに各手法の検出したギャップ数（検出数），およびその中で不正解であったギャップ数（不正解数）を調べた．実験では，これを $R = 1,000$ 回試行し，信頼度 95%（$\alpha = 0.05$）の下での各被覆率ごとのギャップ検出数，不正解数の平均を求めた．

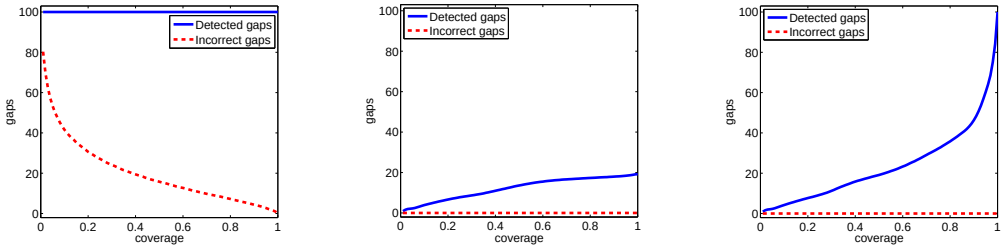
図 1 に近接中心性に対する結果を示す．グラフの横軸（coverage）は被覆率であり，縦軸（gaps）は各試行でのナイーブ法のギャップ検出数が 100 となるように各手法の検出数，不正解数を正規化した値の 1,000 回試行における平均値である．そのため，グラフ中のナイーブ法の検出数は常に 100 となっている．ここで，被覆率 c，r 回目の試行において，各手法が検出したギャップ集合を $A(c,r)$，その中で正しく検出されたギャップ集合を $A^*(c,r)$，ナイーブ法が検出したギャップ集合を $A_{nv}(c,r)$ としたとき，グラフ中の実線（Detected gaps）で表される検出数，および破線（Incorrect gaps）で表される不正解数はそれぞれ次式で定義される．

$$\text{(検出数)} \qquad \frac{1}{R}\sum_{r=1}^{R}\frac{|A(c,r)|}{|A_{nv}(c,r)|} \times 100 \tag{6}$$

$$\text{(不正解数)} \qquad \frac{1}{R}\sum_{r=1}^{R}\frac{|A(c,r) \setminus A^*(c,r)|}{|A_{nv}(c,r)|} \times 100 \tag{7}$$

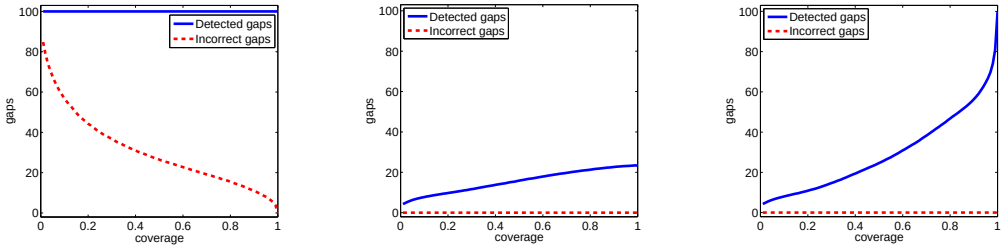
各手法を比較すると，ナイーブ法の検出数はいずれの被覆率でも高いものの，不正解数も多い．被覆率が高くなるにつれて不正解数は減少するが，SE 法，RE 法と比較してその値は非常に大きい．一方，SE 法と RE 法に関しては，被覆率が 0.2 あたりまではほぼ同程度の検出数であり，ナイーブ法と比べると少ないが，SE 法の検出数がその後も大きな伸びを示さないのに対し，RE 法の検出数は徐々に SE 法の検出数を上回り，被覆率が 0.9 を超えるあたりからその数は急増し，最終的には 100 となっている．これは，RE 法が用いるリサンプリング誤差が誤差範囲をより厳密に評価するのに対し，SE 法が用いる標準誤差は誤差範囲を過大評価する傾向にあるためであ

∗2 https://twitter.com



(a) ナイーブ法　(b) SE 法　(c) RE 法

図 1: 近接中心性に対する実験結果



(a) ナイーブ法　(b) SE 法　(c) RE 法

図 2: 媒介中心性に対する実験結果

る．そのため，被覆率が高くなり，中心性の近似値が真の値に近くなっても，SE 法における上位集合の誤差下限と下位集合の誤差上限は多くの場合重複してしまい，ギャップの検出が困難となっている．これに対して RE 法が用いるリサンプリング誤差は誤差範囲を厳密に評価するものの過小評価はしないため，被覆率が高くなるにつれて検出数が多くなる一方，不正解数はほとんど増えず，ほぼ 0 という結果になっている．より安全な誤差範囲を想定する SE 法でも，不正解数に関しては同様の傾向となっている．不正解数が少ないことは，検出したギャップにより高中心性ノード集合が精度よく同定できることを意味する．なお，同じ被覆率では，常に RE 法の検出数は SE 法の検出数以上となっている．図 2 からは，これらの傾向が媒介中心性に対する結果にも共通していることがわかる．

5. おわりに

本稿では，社会ネットワークにおけるノード中心性に関して，サンプリングした一部のノード集合から求められるその近似値のみを用いて，真の中心性の値が高いノード集合を効率的に，かつ高精度で同定する手法を提案した．提案法では，真の中心性の値とその近似値の近似誤差としてリサンプリング誤差を用いることで，与えられた信頼度の下での真の中心性の値の信頼区間を高精度で推定し，中心性の近似値で順序づけられたノード集合を 2 回走査するだけでノード間のすべてのギャップを検出する．実世界の大規模社会ネットワークを用いた評価実験では，信頼度 95% の下，提案法が標準誤差を用いる手法より多くのギャップを検出し，かつ検出したギャップにより高中心性ノード集合を精度よく同定できることを示した．今後の課題としては，他のサンプリングに基づくアプローチとの比較が挙げられる．

謝辞

本研究で用いたデータは東京大学 鳥海不二夫氏，和歌山大学 風間一洋氏によるものである．また，本研究は科学研究費補助金基盤研究 (C) (No. 26330261) の補助を受けた．

参考文献

[Bonacichi 87] Bonacichi, P.: Power and centrality: A family of measures, *Amer. J. Sociol.*, Vol. 92, pp. 1170–1182 (1987)

[Brandes 01] Brandes, U.: A faster algorithm for betweenness centrality, *Journal of Mathematical Sociology*, Vol. 25, pp. 163–177 (2001)

[Brin 98] Brin, S. and L.Page, : The anatomy of a large-scale hypertextual web search engine, *Computer Networks and ISDN Systems*, Vol. 30, pp. 107–117 (1998)

[Chen 13] Chen, W., Lakshmanan, L., and Castillo, C.: Information and influence propagation in social networks, *Synthesis Lectures on Data Management*, Vol. 5(4), pp. 1–177 (2013)

[Freeman 79] Freeman, L.: Centrality in social networks: Conceptual clarification, *Social Networks*, Vol. 1, pp. 215–239 (1979)

[Katz 53] Katz, L.: A new status index derived from sociometric analysis, *Sociometry*, Vol. 18, pp. 39–43 (1953)

[Kleinberg 08] Kleinberg, J.: The convergence of social and technological networks, *Communications of ACM*, Vol. 51, No. 11, pp. 66–72 (2008)

[Newman 01] Newman, M. E. J.: Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality, *Physical Review E*, Vol. 64, p. 016132 (2001)

[Ohara 14] Ohara, K., Saito, K., Kimura, M., and Motoda, H.: Resampling-based Framework for Estimating Node Centrality of Large Social Network, in *Proc. of DS 2014*, pp. 228–239(2014)

[Zhuge 10] Zhuge, H. and Zhang, J.: Topological centrality and its e-Science applications, *Journal of the American Society of Information Science and Technology*, Vol. 61, pp. 1824–1841 (2010)